



Mitigating Data-Imbalance in Regression Problems

Bachelor/ Master Thesis/ HiWi Position (m/w/d)

Background

Machine learning has become increasingly important in scientific applications, for example, in order to build surrogate models predicting the flow field around cars or aircrafts. For these models to be useful, it is required that they not only perform well in the free-stream region far away from the obstacle, but also near the stagnation point, in the attached part of the boundary layer, and in the wake. Generally speaking, a ML model is supposed to make equally accurate predictions in all regions of its training domain. This is, in fact, difficult enough to achieve even if the training data were uniformly distributed over all regimes. However, the training sets are usually also imbalanced, i.e., the number of samples per regime varies significantly. For classification problems, there exist effective strategies to mitigate the effect of data-imbalance. For regression problems, though, it is less obvious how to construct such mitigation strategies.

Project Description

In this thesis, a generalized mitigation strategy, which is applicable and meaningful in classification and regression problems is supposed to be implemented and tested. This method shall further be benchmarked with respect to its effectiveness and computational efficiency against existing methods. To this end, the work is structured as follows:

- Literature review
- Gather a few imbalanced datasets for regression problems
- Setup an automated model development and testing pipeline in Python
- Implement various existing and the proposed generalized imbalance mitigation strategy
- Benchmark all methods
- Documentation of the results

Your Profile

- Solid Python programming skills
- Basic understanding of linear regression

Application

If you are interested, even if you don't fulfill all the listed requirements, please send your application (incl. CV and latest transcript) via [e-mail](#).

Start: as soon as possible

Contact Person:

Hannes Mandler
Institut für Aerodynamik und Gasdynamik
E-Mail: hannes.mandler@iag.uni-stuttgart.de
Tel.: +49 711 685 63461